Joint Feature Distributions for Image Correspondence

Bill Triggs

MOVI, CNRS-INRIA, Grenoble, France

http://www.inrialpes.fr/movi/people/Triggs

Plan

1. Motivation & basic principle
2. Affine JFD's
3. Projective & dual JFD's
4. (If time permits) Tensor joint image.

We need a correspondence model based on ***probability***, not just ***geometry*** !

Potential Advantages

1. More flexible — allow for noise, distortion, small non-rigidities.
2. More efficient & reliable — search is constrained to regions likely to contain the correspondence.
3. Unified approach: one model handles planar, near-planar and 3D cases, no model selection, no near degenerate models to estimate.

Motivation

• Conventional geometric matching constraints are "too idealistic" for practical matching:

1. They only model exact geometry — noise & distortion are "afterthoughts".
2. They assume "all or nothing" — intermediate cases are modelled poorly.

• *E.g.*, owing to depth limitations, correspondences usually lie in *finite well-delimited regions* along epipolar lines

— near-planar scenes, focus / workspace / $\mathbf{p}_\infty$ limits...

• But the epipolar model assumes that correspondences can be *anywhere* along epipolar lines, the homographic one that they are near specific points

— searching the whole line is error-prone and inefficient, searching near a point inadequate.

• Both models are difficult to estimate reliably for near-planar data, and model selection to choose between them is expensive and error-prone.

Joint Feature Distribution Matching

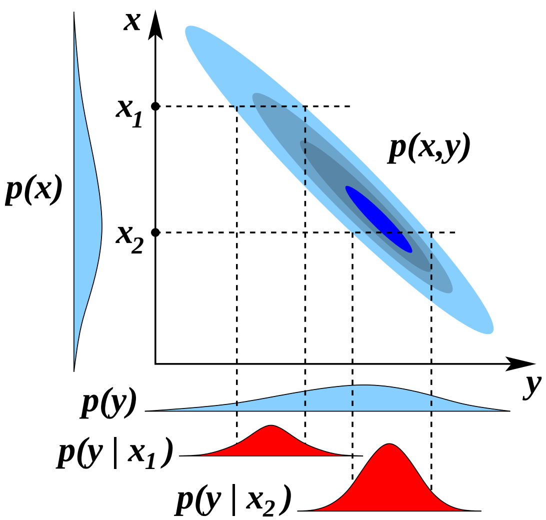
• A natural *probabilistic* matching model, not a rigid geometric one.

1. Estimate **joint probability distribution** for feature population from given correspondences.
2. Use **conditioning** to predict likelihood of match, high-probability search regions for new correspondences...

• **Parametric form** of joint feature distribution should be chosen to characterize *real behaviour of correspondences*

— here, we model geometric matching constraints + noise.

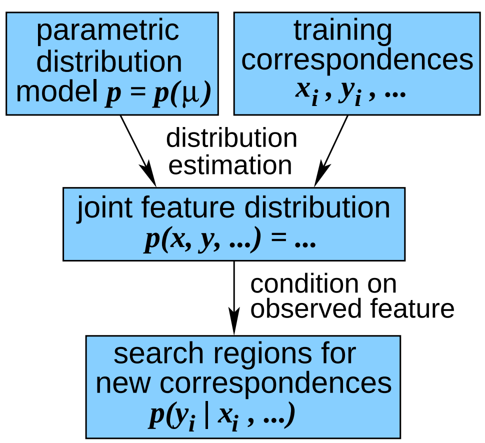
How Joint Feature Distribution Matching Works



Analogies: Geometric *vs*. Probabilistic Matching

Entity	Geometric Matching	Probabilistic Matching
3D camera geometry	Camera projection mapping, matrices $\mathbf{P}_i$	Conditional image feature distributions $\mathbf{p}(\boldsymbol{x}_i \mid \mathbf{f})$
Image signature of camera geometry	Multi-image matching tensors $\mathbf{T}_{ij\ldots k}$	Joint image feature distributions $\mathbf{p}(\boldsymbol{x}, \ldots, \boldsymbol{z})$
Inter-image feature transfer	Tensor based feature transfer $\mathbf{x} \simeq \mathbf{T}_{ij\ldots k} \cdot \mathbf{y} \cdot \ldots \cdot \mathbf{z}$	Conditional image feature distributions $\mathbf{p}(\boldsymbol{x} \mid \boldsymbol{y}, \ldots, \boldsymbol{z})$
Inter-image feature correspondence	Geometric matching constraints $\mathbf{T}_{ij\ldots k} \cdot \mathbf{x} \cdot \ldots \cdot \mathbf{z} = \mathbf{0}$	Probability that features correspond $\mathbf{p}(\boldsymbol{x}, \ldots, \boldsymbol{z})$
Scene reconstruction	Intersection, tensor-based reconstruction methods	Posterior 3D feature probability $\mathbf{p}(\mathbf{f} \mid \boldsymbol{x}, \ldots, \boldsymbol{z})$

JFD Flowchart



Caveats

1 Training & test data must come from same population

• *E.g*. train on near-planar data ⇒ search *only* looks for correspondences near this plane.

• All samples must obey “same noise model”.

2 The relationship to geometry is *partly learned, not fixed in advance*.

• We can choose parametric forms that mimic ordinary matching constraints, but

— they have no notion of an exact underlying tensor satisfying exact consistency relations, *etc*.

— they assume less, so they may not be able to use training correspondences as efficiently as geometric models.

Aside – Homogeneous Covariances

• Many linear / Gaussian calculations can be simplified by using homogeneous coordinates.

• For an uncertain point $\boldsymbol{x}$ with mean $\bar{\boldsymbol{x}}$, variance $\boldsymbol{V}$, homogenize to $\mathbf{x} \equiv \binom{\boldsymbol{x}}{1}$ and define

$$\mathbf{V} \;\equiv\; \langle \mathbf{x}\,\mathbf{x}^\top \rangle \;=\; \begin{pmatrix} \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\top + \boldsymbol{V} & \bar{\boldsymbol{x}} \\ \bar{\boldsymbol{x}}^\top & 1 \end{pmatrix} \qquad \text{homogeneous covariance}$$

$$\boldsymbol{\Lambda} \;=\; \mathbf{V}^{-1} \;=\; \begin{pmatrix} \boldsymbol{V}^{-1} & -\boldsymbol{V}^{-1}\bar{\boldsymbol{x}} \\ -\bar{\boldsymbol{x}}^\top \boldsymbol{V}^{-1} & 1+\bar{\boldsymbol{x}}^\top \boldsymbol{V}^{-1}\bar{\boldsymbol{x}} \end{pmatrix} \qquad \text{homogeneous information}$$

• Means and variances are carried through all calculations together. Also:

$$\chi^2(\boldsymbol{x}\,|\,\boldsymbol{\Lambda}) \;\equiv\; (\boldsymbol{x}-\bar{\boldsymbol{x}})^\top \boldsymbol{V}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}) \;=\; \mathbf{x}^\top \boldsymbol{\Lambda}\,\mathbf{x} - 1$$

Homogeneous Scatter Matrix

• The **homogeneous scatter matrix** of a set of training points $\{\boldsymbol{x}_i\}_{i=1\ldots n}$ is:

$$\mathbf{V} \;\equiv\; \tfrac{1}{n}\textstyle\sum_i \mathbf{x}_i\,\mathbf{x}_i^\top$$

• If the point population is modelled as a Gaussian, the ML estimates of its homogeneous covariance & information are $\mathbf{V}$ and $\boldsymbol{\Lambda} = \mathbf{V}^{-1}$.

• If there are only a few points and $\boldsymbol{x}_i$ is uncertain with noise $\boldsymbol{V}_i$, the **smoothed scatter matrix** gives smoother estimates of the population covariance:

$$\mathbf{V} \;\equiv\; \frac{1}{n}\sum_i \langle \mathbf{x}_i\,\mathbf{x}_i^\top \rangle \;=\; \frac{1}{n}\sum_i \begin{pmatrix} \bar{\boldsymbol{x}}_i\bar{\boldsymbol{x}}_i^\top + \boldsymbol{V}_i & \bar{\boldsymbol{x}}_i \\ \bar{\boldsymbol{x}}_i^\top & 1 \end{pmatrix}$$

The Affine Joint Feature Distribution

• Gaussians are the natural JFD model for affine cameras, as the matching constraints are linear.

Estimating an affine JFD

1. Form $(2m+1)$-D **affine joint image** vectors from the coordinates of corresponding features: $\mathsf{x}_i \;\equiv\; (\boldsymbol{x}_i, \ldots, \boldsymbol{z}_i, 1)^\top$.

2. Form the homogeneous scatter matrix of the training data $\mathsf{V} = \frac{1}{n}\sum_i \mathsf{x}_i\,\mathsf{x}_i^\top$.

3. Model the joint feature population as a Gaussian with homogeneous covariance V, information $\boldsymbol{\Lambda} \;=\; \mathsf{V}^{-1}$.

Using an affine JFD

• Given the JFD $\mathrm{p}(\boldsymbol{x}, \ldots, \boldsymbol{z})$ and values for some of its variables $\boldsymbol{y}, \ldots, \boldsymbol{z}$, the conditional distribution $\mathrm{p}(\boldsymbol{x}, \ldots \,|\, \boldsymbol{y}, \ldots, \boldsymbol{z})$ defines high-probability search regions for the remaining variables $\boldsymbol{x}, \ldots$

• The affine JFD and its conditionals are Gaussian, so familiar Gaussian-model formulae reappear — *c.f.* linear regression, Kalman filtering.

• *E.g.*, freezing variables $\boldsymbol{y}$ in $\mathsf{x} = (\boldsymbol{x}, \boldsymbol{y}, 1)^\top$ and rewriting $1 + \chi^2(\mathsf{x}\,|\,\boldsymbol{\Lambda}) \;=\; \mathsf{x}^\top \boldsymbol{\Lambda}\,\mathsf{x}$ in terms of the remaining ones $\boldsymbol{x}$ gives a conditional Gaussian with:

$$\begin{aligned} \text{mean} \;&=\; \bar{\boldsymbol{x}} - (\boldsymbol{\Lambda}_{\boldsymbol{xx}})^{-1}\,\boldsymbol{\Lambda}_{\boldsymbol{xy}}\,(\boldsymbol{y}-\bar{\boldsymbol{y}}) \\ \text{variance} \;&=\; (\boldsymbol{\Lambda}_{\boldsymbol{xx}})^{-1} \end{aligned}$$

Subspaces and Noise

• For $m \geq 2$ noiseless affine images, the scatter V and its JFD lie in a 4D subspace — the "affine joint image".

• But we do *not* enforce this here as we want to characterize the real behaviour of correspondences, not reproduce an ideal geometric model !

• Noise, distortion and independent motions all increase the subspace dimension, *c.f*.

— Boult-Brown & Costeira-Kanade factorization

— Shashua's extended projections

Projective Joint Feature Distributions

• Projective JFD's are similar to affine ones but use **tensor product coordinates** not direct sums.

• ***Rationale:*** As with geometric matching constraints:

— Tractable linear / Gaussian models only handle linear constraints

— Projective matching constraints are multilinear but not linear

$\Rightarrow$ quasi-linearize the problem using tensored coordinates.

• *E.g*. affine epipolar models use **affine joint image** coordinates:

$$\boxed{(\mathbf{x}, \mathbf{x}', 1) \;=\; (x,\; y,\; x',\; y',\; 1)}$$

While projective epipolar models use **tensor joint image** coordinates:

$$\mathbf{x} \otimes \mathbf{x}' \;=\; (\underbrace{xx',\; xy',\; yx',\; yy'}_{\text{nonlinear "perspective coordinates"}}, \underbrace{x,\; y,\; x',\; y',\; 1}_{\text{affine coordinates}})$$

Smoothing Strategies

• Each correspondence gives a rank-1 contribution to the scatter matrix V, so we need $n \geq 2m+1$ correspondences to make V nonsingular

— four specify the affine joint image, the rest the noise behaviour.

• To remove the singularity of V for small n, we add a small regularizer $\epsilon^2 \cdot$ diag$(1, ..., 1, 0)$ to V before inverting

— in general this gives an over-focused JFD...

— but low-dimensional conditionals are usually still correct for $n \gg 4$.

• Alternatively, we can regularize using smoothed scatter matrices, at the risk of double-counting some of the noise.

Form of Projective JFD

• We model the projective joint feature distribution as a Gaussian $\mathbf{p}(\mathbf{x}, ..., \mathbf{z}) \sim e^{-L(\mathbf{x},...,\mathbf{z})/2}$ in the tensor product coordinates, where:

$$\begin{aligned} L(\mathbf{x}, ..., \mathbf{z}) \;&\simeq\; (\mathbf{x} \otimes ... \otimes \mathbf{z})^\top \,\mathbf{\Lambda}\, (\mathbf{x} \otimes ... \otimes \mathbf{z}) \\ &=\; \mathbf{\Lambda}_{A...D\; A'...D'} \cdot (\mathbf{x}^A \mathbf{x}^{A'}) \cdot ... \cdot (\mathbf{z}^D \mathbf{z}^{D'}) \end{aligned}$$

• The JFD model can represent *arbitrary weighted sums of algebraic errors*:

$$L(\mathbf{x}, ..., \mathbf{z}) \;=\; \sum_k \lambda_k \, |\mathbf{t}^k_{A...D}\, (\mathbf{x}^A ... \mathbf{z}^D)|^2 \;=\; (\mathbf{x} \otimes ... \otimes \mathbf{z})^\top \mathbf{\Lambda} (\mathbf{x} \otimes ... \otimes \mathbf{z})$$

$$\mathbf{\Lambda}_{A...D\; A'...D'} \;\equiv\; \sum_k \lambda_k \, \mathbf{t}^k_{A...D} \, \mathbf{t}^k_{A'...D'}$$

e.g., these might represent (components of) geometric matching constraints.

• Conditioning this JFD is trivial: freezing the values of some variables gives a model of the same form in the remaining ones.

• Conditioning down to a single image gives a standard Gaussian.

Estimating the Projective JFD

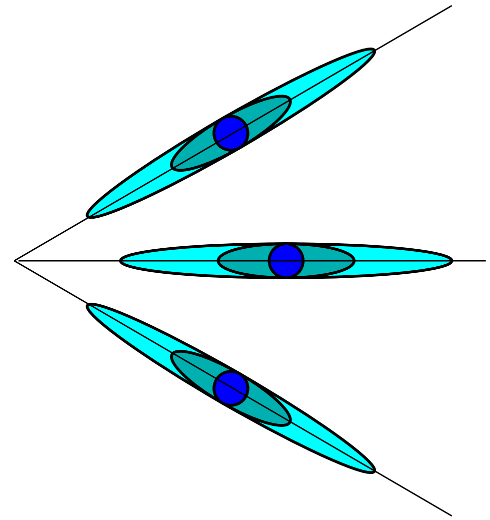
1 Build the **homogeneous scatter tensor**

$$\mathsf{V} \;=\; \sum_i (\mathbf{x}_i \otimes \ldots \otimes \mathbf{z}_i) \otimes (\mathbf{x}_i \otimes \ldots \otimes \mathbf{z}_i)$$

- This is linear in $\mathbf{x}_i\,\mathbf{x}_i^\top$, so we can also use smoothed scatter matrices here.

2 Write V as a $3^m \times 3^m$ matrix, regularize by adding $\epsilon^2\mathrm{diag}(1, ..., 1, 0)$ as before, and invert to find $\mathbf{\Lambda}$.

- *E.g.*, for 2 images V is $\mathsf{A}^\top\,\mathsf{A}$ where A is the usual $n \times 9$ fundamental matrix data matrix with rows $(xx', xy', yx', yy', x, y, x', y', 1)$.



Projective JFD *vs*. Fundamental Matrix

- Write $\mathbf{F}$ as a 9-vector f. The epipolar error can be written as a JFD residual:

$$|\mathbf{x}^\top\,\mathbf{F}\,\mathbf{y}|^2 \;=\; (\mathbf{x} \otimes \mathbf{y})^\top\,\mathbf{\Lambda}\,(\mathbf{x} \otimes \mathbf{y}) \qquad \text{where} \;\; \mathbf{\Lambda} \;\equiv\; \mathsf{f}\,\mathsf{f}^\top$$

- As f is estimated as the smallest eigenvector of $\mathsf{V} \;=\; \mathsf{A}^\top\,\mathsf{A}$, converting the JFD model to the epipolar one amounts to truncating the information $\mathbf{\Lambda} = \mathsf{V}^{-1}$ at its largest eigenvector $\mathsf{f}\,\mathsf{f}^\top$.

- In general, $\mathbf{\Lambda}$ represents quadratic constraints on $\mathbf{x} \otimes \mathbf{y}$, and conditioning on $\mathbf{x}$ gives an elliptical search region for $\mathbf{y}$. The ellipticity depends mainly on the ratio λ_1/λ_2 of the largest and second largest eigenvalues of $\mathbf{\Lambda}$:

$\lambda_1/\lambda_2 \to \infty$	An ideal fundamental matrix, search an infinite strip along the epipolar line.
$\lambda_1/\lambda_2 \to 1$	An ideal homography, search a circular region.
in between	Intermediate geometry, search an ellipse aligned with the local epipolar line.

Limitations of Standard Projective JFD Model

- The above projective JFD model works well in the practically most important case of 2 images, but it has several limitations:

1. Inexactness

- The model is not quite Gaussian as the 3^m components of the rank-1 tensor $(\mathbf{x} \otimes \ldots \otimes \mathbf{z})$ are not independent variables
- The true model is a Gaussian restricted to this rank 1 subvariety.
- In principle this invalidates the estimation rule $\mathbf{\Lambda} \;=\; \mathsf{V}^{-1}$, which is based on the normalization an unconstrained Gaussian

  — the correct rule is unknown as the required integral is intractable.

- In practice we still use $\mathbf{\Lambda} \;=\; \mathsf{V}^{-1}$. It is exact for affine images, and so far it seems to work well in other cases.

2. Speed

- $3^m \times 3^m$ matrices must be handled.
- This is slow for large m, but not easily fixable.

3. Inefficient use of correspondences

- The projective JFD uses training correspondences inefficiently for $m \geq 3$.
- For $m = 2, 3, 4...$ images, $n = 8, 17, 31...$ correspondences are needed just to estimate the geometric subspaces, or $6, 10, 15, ...$ for coplanar points
- The extra correspondences are used to learn the behaviour of the nonlinear "perspective coordinates" that were added to the model, beyond the original $3m+1$ affine ones
- As in the affine case, characterizing the noise requires yet more correspondences.
- The inefficiency occurs because each correspondence gives just one constraint on the model (all contributions to V are rank 1).
- To fix this we use higher-rank *dual covariances* in models that mimic the index structure of point-on-line (homographic, trifocal...) matching constraints

Dual Covariances

- Just as standard homogeneous covariances describe the positions and uncertainties of points, **dual covariances** describe the positions and uncertainties of the lines (hyperplanes) through them.
- The dual covariance of a fixed point $\mathbf{x}$ is the scatter matrix of "the uniform distribution" of lines through it, which we *define* to be:

$$\widetilde{\mathbf{V}}_{AA'} \equiv [\mathbf{x}]_\times^\top \, \mathbf{I} \, [\mathbf{x}]_\times$$

The quadric $\mathbf{I}^{AA'}$ defines the notion of "uniformity"

— usually, we choose $\mathbf{I} = \mathbf{I}_{3\times3}$ in the current frame.

- For uncertain points, the dual distributions get blurred, but we can still swap between normal and dual covariances without losing information:

$$\mathbf{V}^{AA'} = \left\langle \mathbf{x}\,\mathbf{x}^\top \right\rangle = \begin{pmatrix} x^2+\mathbf{V}_{xx} & xy+\mathbf{V}_{xy} & x \\ xy+\mathbf{V}_{xy} & y^2+\mathbf{V}_{yy} & y \\ x & y & 1 \end{pmatrix} \qquad \text{point covariance}$$

$$\begin{aligned} \widetilde{\mathbf{V}}_{AA'} &= \left\langle [\mathbf{x}]_\times^\top \, \mathbf{I} \, [\mathbf{x}]_\times \right\rangle \\ &= \begin{pmatrix} 1+y^2+\mathbf{V}_{yy} & -xy-\mathbf{V}_{xy} & -x \\ -xy-\mathbf{V}_{xy} & 1+x^2+\mathbf{V}_{xx} & -y \\ -x & -y & x^2+y^2+\mathbf{V}_{xx}+\mathbf{V}_{yy} \end{pmatrix} \qquad \text{dual covariance} \\ &= \mathbf{u}\,\mathbf{u}^\top + \mathbf{v}\,\mathbf{v}^\top + \mathbf{w}\,\mathbf{w}^\top \qquad (\mathbf{u}, \mathbf{v}, \mathbf{w}) = \begin{pmatrix} 0 & -1 & y \\ 1 & 0 & -x \\ -y & x & 0 \end{pmatrix} \end{aligned}$$

Point-on-Line JFD's

- Given any point-on-line matching constraint (homography, trifocal...), we can use dual covariances to build a corresponding projective JFD model.
- Models containing duals are:

— less symmetric between the images

— less general: geometric assumptions have been built in in the sense that the constraint must hold for any line through the point

— more efficient: each independent (combination of) line(s) gives a new constraint on the model.

Example — homographic JFD model

- The error of the homographic model $\mathbf{y} \simeq \mathbf{H}\,\mathbf{x}$ can be written:

$$\boxed{\left\|[\mathbf{y}]_\times\,\mathbf{H}\,\mathbf{x}\right\|^2 \;=\; (\mathbf{H}^A_B\,\mathbf{H}^{A'}_{B'})(\widetilde{\mathbf{V}}^{BB'}_{\mathbf{y}})(\mathbf{x}^A\,\mathbf{x}^{A'})}$$

- This model is analogous to the epipolar JFD with $\mathbf{F} \to \mathbf{H}$ and $\mathbf{V}_\mathbf{y} \to \widetilde{\mathbf{V}}_\mathbf{y}$, but as $\widetilde{\mathbf{V}}_\mathbf{y}$ has rank ≥ 2 we get a stronger correspondence constraint that needs only 4 correspondences.
- Like standard homographies, the homographic JFD model basically assumes coplanar data. But it also models uncertainties and degeneracies within the plane.
- All other point-on-line constraints are similar:

— write the constraint in terms of $[\,\mathbf{x}\,]_\times$

— square, and replace factors of $[\,\mathbf{x}\,]_\times^\top\;\mathbf{I}\;[\,\mathbf{x}\,]_\times$ with $\widetilde{\mathbf{V}}_\mathbf{x}$.

Summary of Joint Feature Distribution Approach

- JFD's are the natural *probabilistic* (not geometric !) model for feature correspondence

— they gracefully handle near-degenerate geometry, distortion, non-rigidity...

1. Choose a parametric form for JFD that can reproduce geometric matching constraints

— Gaussian for affine images

— quasi-Gaussian in tensored coordinates for projective ones

— dual forms use training data more efficiently for $m > 2$ projective images.

2. Estimate the feature population JFD $\mathbf{p}(\boldsymbol{x}, ..., \boldsymbol{z})$ from training correspondences $(\boldsymbol{x}_i, ..., \boldsymbol{z}_i)$.
3. Predict search regions for further correspondences by conditioning: $\mathbf{p}(\boldsymbol{x}\,|\,\boldsymbol{y}, ..$

Future work on JFD's

1. Develop stabler numerical methods: inversion $\to$ QR.
2. Test on real images — robustification and RANSAC?
3. Try to replace algebraic error weightings with more statistical ones.

Tensor Joint Image

- A new algebraic geometry approach for studying multi-image projection and matching constraints

$$
\begin{tikzcd}[column sep=huge, row sep=large]
\mathbb{P}^{a} \arrow[r, "\mathbf{X}^a \to \mathbf{X}^a \cdot \ldots \cdot \mathbf{X}^d", "\text{Veronese mapping}"'] \arrow[d, "\text{joint projection}", "\begin{array}{c}\mathbf{X}^a \\ \downarrow \\ (\mathbf{P}_a^A\,\mathbf{X}^a, \ldots, \mathbf{P}_a^D\,\mathbf{X}^a)\end{array}"'] & \mathbb{P}^{(a\ldots d)} \arrow[d, "\begin{array}{c}\mathbf{T}^{a\ldots d} \\ \downarrow \\ \mathbf{P}_a^A \ldots \mathbf{P}_d^D\,\mathbf{T}^{a\ldots d}\end{array}", "\text{tensor projection}"'] \\
\mathbb{P}^{A} \times \ldots \times \mathbb{P}^{D} \arrow[r, "\text{Segré mapping}", "(\mathbf{x}^A, \ldots, \mathbf{z}^D) \to \mathbf{x}^A \cdot \ldots \cdot \mathbf{z}^D"'] & \mathbb{P}^{A\ldots D}
\end{tikzcd}
$$

- The Segré mapping quasi-linearizes the matching constraints in exchange for using a nonlinear coordinate system.
- The matching constraints are exactly the linear forms on $\mathbb{P}^{A\ldots D}$ that contain the **tensor joint image** — the image of $\mathbb{P}^{a}$.

Future work on tensor joint image

- Study the power series expansion

$$\boxed{\text{projective case} \;=\; \text{affine case} \;+\; \text{“perspective” corrections}}$$

Here, the “perspective corrections” are due to the known curvature of Segré embedding, *not* problem specific geometry.